Credit Scoring and Disparate Impact

Elaine Fortowsky & Michael LaCour-Little

Wells Fargo Home Mortgage

7911 Forsyth Blvd, Suite 600

Clayton, MO 63105

Dec 13, 2001

The views expressed are the authors alone and not necessarily those of Wells Fargo or any of its affiliates

Abstract

We analyze the problem of disparate impact in credit scoring and evaluate three approaches to identifying and correcting the problem, namely: 1) post-development univariate test with variable elimination, 2) post-development multivariate test with variable elimination, 3) control variable approach with coefficient adjustment. The third approach is a new innovation developed by the authors. Results are illustrated with simulation data calibrated to actual distributions of typical variables used in score development. Results show that controlling disparate impact by eliminating variables may have unintended and undesirable consequences.

Key words: credit scoring, discrimination, disparate impact, ECOA.

1 Introduction

Credit scoring, that is, statistically based evaluation of borrower credit quality as part of the loan underwriting process, has been a part of consumer lending for decades. Fair, Isaac, and Company (FICO) is usually credited with developing the first credit scoring systems for retail chains offering consumer credit during the 1950s and continues to be the industry leader. Scoring has become much more prevalent during the past decade, especially in home mortgage lending, since the government-sponsored enterprises, Fannie Mae and Freddie Mac, adopted their use in 1995. Credit card issuers have been using credit scoring systems for decades and the explosion of pre-authorized credit offers during the mid-1990s is generally thought attributable to creditor data mining for eligible households based on credit score thresholds or specific credit bureau attributes. For a review of issues in the use and development of credit scores, see Mays (2001). For an industry perspective on the use of credit scoring with low-to-moderate income or high-minority populations, see Martell, et al [1999]. For current academic work on the effect of various scoring approaches on low-income households, see Collins et al [2001].

Allegations of discrimination in financial services have also increased over the past decade, particularly in the area of home mortgage lending. In the mortgage market, these charges arise from two uncontested empirical observations. First, the volume of home mortgage loans per mortgagable dwelling unit in predominantly white areas is two to three times the rate in minority neighborhoods; second, the rejection rate for minority mortgage loan applicants is roughly twice that of white applicants (Fix, Galster, and Struyk [1993]). For competing views on the controversy over discrimination in mortgage lending, see Ladd [1998] and LaCour-Little [1999]. More recently, the auto finance industry has been subject to a series of class action lawsuits that allege discriminatory pricing and disparate impact on minority groups. These charges arise from the pricing practices of auto dealers and captive finance companies, which allegedly result in African-Americans paying more for auto credit than similarly situated white borrowers (see U.S. Department of Justice [2000] for summary of facts and issues in a representative case).

Institutional lenders, especially those affiliated with depository institutions, are subject to an elaborate regulatory regime. Regulatory agencies conduct periodic examinations to ensure that credit scoring, if used, is statistically sound and consistently applied to produce fair outcomes for all loan applicants. The major regulatory agencies coordinate procedures through the Federal Financial Institutions Examination Council (FFIEC). For a detailed description of these procedures, which sometimes include estimation of regression equations, see FFIEC [1999]. These activities are intended to ensure compliance with fair lending laws, especially the Equal Credit Opportunity Act (ECOA).

ECOA was first enacted in 1974 and is the primary statute under which legal proceedings alleging discrimination in financial services are brought. Originally designed to protect women from differential treatment, the law was amended to include race and other protected categories in 1976. ECOA regulates all types of credit, not just mortgage lending and prohibits discrimination based on race or color, religion, national origin, sex, marital status, age, receipt of public assistance, or good faith exercise of rights under the Consumer Credit Protection Act.

Economists have long been interested in explaining the phenomenon of discrimination since the seminal work of Becker [1959]. Becker developed the concept of "a taste for discrimination", which implies that firms must forego profits, customers, or workers in order to indulge their discriminatory preferences (prejudice). Markets punish such behavior; hence, Becker's theory implies a decline over time in discrimination. An alternative theory based on the concept of statistical discrimination" is associated with both Phelps [1972] and Arrow [1973]. In the Phelps-Arrow sense, discriminating individuals or firms take race or gender as a cheap signal of other hard-to-observe characteristics that may be correlated with outcomes they wish to control. Hence if females are less productive than males, on average, in a particular job, then it may be rational, in some sense, for firms to prefer males to females and not invest the extra effort necessary to discover the true expected productivity of female job applicants. Analogously, if minority borrowers default more frequently, it may be rational for lenders to prefer white borrowers, especially if the cost of discovering true expected loan performance is relatively high. Unfortunately, market discipline cannot be expected to eradicate statistical discrimination.

We now turn to the legal theory of discrimination. In this context, discrimination may be most simply defined as differential treatment of otherwise similarly situated individuals on the basis of race, gender, national origin, or other protected characteristic. The general legal theory under which discrimination is unlawful is the equal protection provision of the Fourteenth Amendment. Courts have held that legal classifications based on race or gender are suspect and may be sustained only where a compelling interest can be shown (Kaye [1986]). Racial discrimination by government agencies themselves, either in intent or effect, has been alleged in criminal prosecution, capital sentencing, jury selection, and awarding of government entitlements, as well as many other areas.

Discrimination in employment is explicitly prohibited by Title VII of the Civil Rights Act of 1964. Title VII prohibits discrimination, both disparate treatment and disparate impact, on the basis of race, national origin, sex, and religion. Disparate treatment is viewed as intentional, in the sense that the discriminating party took into account, overtly or covertly, the prohibited characteristic of the victim of discrimination. Claims of disparate impact, on the other hand, do not require a showing of intentional discrimination. Moreover, a showing of business necessity may rebut claims of disparate impact. Business necessity means that the factor used to discriminate, typically a positive correlate of the prohibited factor, serves a valid business purpose and is not merely a pretext for overt discrimination based on the prohibited factor . We will reserve discussion of the business necessity argument for later research.

In the lending context, discrimination may consist of either (1) refusing to transact or (2) varying the terms of the transaction. Discrimination is generally categorized into three types: (1) overt discrimination, (2) disparate treatment, and (3) disparate, or adverse, impact. Overt discrimination occurs when a lender openly discriminates based on a prohibited factor. Disparate treatment occurs when lenders treat applicants differently based on a prohibited factor. Adverse impact occurs when a business practice is applied uniformly, but has a disparate impact on a protected class.

Much of the law on discrimination in lending evolved out of employment law (Siskin [1995]). But case law is firmly established in employment discrimination, while few lending discrimination actions are rarely adjudicated. Complaints are often resolved by consent decree, in which the lender admits no wrongdoing but commits to revise its policies and practices, occasionally compensating victims of the actions alleged to be discriminatory. Statistical evidence intended to show a pattern or practice of discriminatory behavior is often developed, particularly for class-action cases.

Sandler and Biran [1995] critique the use of statistics in legal proceedings involving complaints of discrimination, based mainly on employment discrimination cases. Courts have taken the position that statistical analysis may be offered as evidence if (1) the model is reasonably well specified, (2) statistics are based on a "proper pool" of applicants, (3) statistics show "substantial" discriminatory effect, and (4) the analysis can isolate the effects of a particular criterion used in the decision-making process . The first criterion requires the set of independent variables in any regression to be reasonably comprehensive, i.e., with little likelihood of substantial omitted variable bias. The second requires that the data sample be of adequate size. The third requires that the measure of discriminatory effect must be large, for example, the coefficient on the race variable must be of significant magnitude relative to other variables and have a significant t-statistic . The fourth requires that differential denial rates be attributable to a specific variable not a combination of many factors. In summary, limited case law makes definitive statements about lending discrimination cases difficult, although analogies to employment law offer some guidance. If we assume that overt discrimination is rare, the distinction between disparate treatment and disparate impact is essential: disparate treatment constitutes intentional discrimination and cannot be rebutted, whereas disparate effect is unintentional and may be rebutted by a business necessity argument. Most academic and policy-oriented research to date has focused on disparate treatment, as in the well-known study by Munnell et al [1996].

In this paper, we extend the discussion to the topic of disparate impact and consider how regulatory agencies might test for its existence and measure its magnitude, as well as how score developers might modify procedures to minimize its effect. Our work is not directed at any particular segment of consumer lending, though most of our experience is in the mortgage arena. Accordingly, the default rates and inter-group differences we present should be viewed as broadly representative and not as examples of any particular line of the consumer lending business.

The plan for the balance of the paper is as follows. In the next (second) section we provide an overview of our approach. In the third section, we describe the data generating process and credit score construction, and develop a measure of disparate impact. In the fourth section, we present and describe the pseudo data generated by our simulation process. In the fifth section, we estimate scorecards. In the sixth section, we present results of univariate and multivariate tests and discuss results. In the seventh section, we present an alternative method and assess its benefits. The final section discusses issues related to implementation given ECOA and offers concluding comments.

2 overview

Disparate impact analysis of scorecards has to date proceeded without a rigorous definition of what the testing, or the corrective action, is meant to accomplish.

We start by defining the criteria that we would like a 'good' disparate impact test to satisfy, and the criteria that we would like a 'good' corrective action plan to satisfy. After defining our performance criteria, we discuss the sample and the estimated scorecards that will be the subject of our disparate impact analysis. We then evaluate two disparate impact testing procedures, univariate and multivariate, and show that the multivariate is superior. Following this we evaluate the standard corrective action (which has been used with both the univariate and the multivariate test) of eliminating variables, and show that it does not actually correct disparate impact. We then present a new corrective action approach that does eliminate disparate impact, namely, retaining the original set of variables but adjusting the scorecard weights by employing minority status as a control variable during estimation.

3 Performance Criteria and Data Generation

Consider the default process

$$Y_i^* = \boldsymbol{\beta}' \boldsymbol{x}_i + \varepsilon_i; \ \varepsilon_i \sim logistic(\mu_i, \sigma^2) \tag{1}$$

$$\mu_i = \begin{cases} \mu_P \geq 0, & \text{if individual } i \text{ is in the protected class;} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

$$Y = \begin{cases} 1, & \text{if } Y^* >= c; \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

where the ε_i are assumed to be independent and $Y = 1$ indicates that the loan has defaulted. In this paper we will consider two variations on the default process. In the first one, Process 1, we will set $\mu_P > 0$. The error term will then be centered around a larger mean for the protected class than for the non-protected class, so the protected class will have a higher default rate, all else being equal. This construction allows disparate impact to be possible (a variable coefficient can act as a proxy for protected class status only if protected class status actually helps predict credit-worthiness). In the second variation, Process 2, we will set $\mu_P = 0$. The error term will then be centered around zero for both classes and disparate impact will not arise.

Define the score of observation i as

$$S_i = \boldsymbol{\beta}' \boldsymbol{x}_i \tag{4}$$

We will refer to the above as the class-neutral score. Let the observations associated with the highest 5% of scores form the reject set, and let the group consisting of protected class members from this set be denoted by R^P.

Now let $\hat{\boldsymbol{\beta}}$ be the maximum likelihood estimate of the scaled coefficient vector $\frac{1}{\sigma}\boldsymbol{\beta}$ and define the estimated score of observation i as

$$\hat{S}_i = \hat{\boldsymbol{\beta}}' \boldsymbol{x}_i \tag{5}$$

Let the observations associated with the highest 5% of scores form the reject set, and let the group consisting of protected class members from this set be denoted by $\hat{R}^P$. Our measure of disparate impact is

$$\Psi = \frac{\Sigma I((i \in \hat{R}^P) \cap (i \notin R^P))}{\Sigma I(i \in \hat{R}^P)} \tag{6}$$

which is the percentage of rejected protected-class members that would not have been rejected using the class-neutral scores. A 'good' disparate impact test is one that differentiates between $\Psi > 0$ and $\Psi = 0$, and a 'good' corrective action plan is one that achieves $\Psi = 0$ (or at least close to zero).

The disparate impact tests and corrective actions will be evaluated on a simplified scorecard consisting of just three variables: FICO[1], debt ratio, and income. The first two factors are widely viewed as important predictors of credit risk, whereas income is a more questionable factor, due to its correlation with protected group status. Due to privacy concerns, we will not estimate the scorecard on actual data; instead, we will generate a simulated sample in which the three variables are assumed to be jointly normally distributed. The means and variance/covariances of the three variables will be based on their actual empirical distribution within a random group of borrowers who applied for mortgage loans in the first half of 2001, with separate distributions for the protected class and the non-protected class. The reader may wish to think of the protected class as members of racial minority groups; however, the same approach applies to gender, age, or any other category protected from discrimination under the law. Our use of actual empirical distributions to calibrate the data-generating process is meant to ensure that our results pertain to the typical types of data that a credit-scorecard builder might face.

For consistency with later estimation technique, the error term is assumed

[1]FICO is a generic credit score marketed by *Fair, Isaac.* It takes on values from 300 to 900 with every 20 point decrease doubling the odds of default.

to be logistically distributed. We did not attempt to get empirical evidence on the paramters of the error distribution, partly to expedite this research and partly because at best we would only have evidence for the accepted applications whereas our scorecard is meant to be applied to all applications, both accepted and rejected.

We will now present the empirical parameters for our data-generating process. It is assumed that the vector of regressors is ordered as:

$$x = \begin{pmatrix} FICO \\ RATIO \\ INCOME \end{pmatrix} \tag{7}$$

For the non-protected class the parameters are

$$\mu_P = \begin{pmatrix} 712 \\ 37 \\ 83 \end{pmatrix}, \sigma_P^2 = \begin{pmatrix} 4030 & -118 & 575 \\ -118 & 139 & -261 \\ 575 & -261 & 10618 \end{pmatrix} \tag{8}$$

where income is measured in '000s, while for the protected class they are

$$\mu_{NP} = \begin{pmatrix} 668 \\ 39 \\ 58 \end{pmatrix}, \sigma_{NP}^2 = \begin{pmatrix} 4677 & -52 & 374 \\ -52 & 123 & -93 \\ 374 & -93 & 2654 \end{pmatrix} \tag{9}$$

The error term is distributed as

$$\varepsilon_i \sim logistic(\mu_i, \sigma = 5,000) \tag{10}$$

with Process 1 defined by

$$\mu_i = \begin{cases} 400, & \text{if individual } i \text{ is in the protected class;} \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

and Process 2 defined by

$$\mu_i = 0 \tag{12}$$

Loan defaults are generated by:

$$Y_i^* = 300 - FICO_i + 20 \times RATIO_i + \varepsilon_i \tag{13}$$

$$Y_i = \begin{cases} 1, & \text{if } Y_i^* >= 700; \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

where $Y_i = 1$ indicates that the loan has defaulted.

Note that income does not appear in the default equation. It was intentionally left out to provide an additional performance hurdle for the tests and corrective procedures. We created the sample for Process 2 first, using (8) thru (10) and (12) thru (14) to generate 10,000 observations for the protected class and 80,000 observations for the non-protected class. Income and debt ratio were truncated at zero. To get the sample for Process 1, we took the protected class observations from the Process 2 sample and added 400 to each error term. We then used (13) and (14) to generate defaults. The samples for Process 1 and Process 2 are thus identical except for the error term and the default indicator on the protected class observations. Descriptive statistics for the shared variables from the samples are presented in Figure 1.

Figure 1: Descriptive Statistics

Variable	Protected Class Mean	Non-Protected Class Mean
Fico	667	712
Debt Ratio	39%	37%
Income	$61,000	$95,000

4 Estimated Scorecards

To develop our scorecards, the simulated data was used to estimate the logit model:

$$\ln\left(\frac{P(Y)}{1-P(Y)}\right) = \frac{1}{\sigma}\alpha + \frac{1}{\sigma}\beta_1 \times FICO + \frac{1}{\sigma}\beta_2 \times RATIO + \frac{1}{\sigma}\beta_3 \times INCOME \quad (15)$$

As noted earlier, we include income in the model specification even though it does not impact the data-generating process. This typifies actual modeling, in which the statistician does not know *a priori* exactly what set of explanatory variables is generating the target variable, and will very likely include some irrelevant ones in the initial specification.

We estimated the model using *proc logistic* in *SAS*. The results are presented in Figure 2. In the Process-1 model all variables are significant at the level $\alpha < 0.0001$ based on the Wald Chi-Square statistic. In the Process-2 model, income is not significant. The change in significance of the income variable as we move from Process 1 to Process 2 is due to the fact that it is correlated with the error term in Process 1 but not in Process 2.

Figure 2: Estimated Equations

	— Process 1 —		— Process 2 —	
Variable	Estimated Coefficient	Significance	Estimated Coefficient	Significance
Intercept	1.880	<.0001	-10.487	<.0001
Fico	-0.016	<.0001	-0.026	<.0001
Debt Ratio	0.176	<.0001	0.519	<.0001
Income (00,000)	-0.336	<.0001	-0.030	<.3439

To derive scorecards from Figure 2 we transformed the coefficients so that the predicted values using the new coefficients ranged between 0 and 100. For the Process-1 model this required setting the intercept to 70 and multiplying each of the regressor coefficients by $2.5/\ln(2)$, while for the Process-2 model it required setting the intercept to 50 and multiplying each of the regressor coefficients by $1.2/\ln(2)$. Such transformations are standard practice, and speak to the fact that only the ordinal ranking of the scores is important. Note that the predicted values from the transformed models have the property that the odds of default double for every 2.5 (1.2) points in the Process-1 (Process-2) model. For clarity we will refer to the predicted values from the transformed models as transformed scores, which are distinct from the raw predicted values used in formula (6)

Scorecards are presented in Figure 3. The last row gives the disparate impact measure for each scorecard based on formula (6). For Scorecard 1, 25.6% of the rejected protected class applicants would not have been rejected by the class-neutral scores. For Scorecard 2 the number is 0.1% (effectively zero).

Figure 3: Scorecards for Process 1 and Process 2

	Scorecard 1 *(Process 1)*	Scorecard 2 *(Process 2)*
	– Point Weights –	
Base Points	70	50
Fico	-0.058	-0.045
Debt Ratio	0.635	0.899
Income (00,000)	-1.212	-0.052
	Ψ=25.6	Ψ=0.1

We will now demonstrate how the univariate and multivariate tests are applied to our scorecards. For a test to be 'good', we want it to identify the fact that Scorecard 1 is characterized by $\Psi > 0$ and Scorecard 2 is characterized by $\Psi = 0$. We will see that the multivariate test correctly identifies both cases while the univariate test erroneously concludes that $\Psi > 0$ for Scorecard 2.

5 Tests and Results

The univariate disparate impact test is based on tables such as 4. The column labelled *Point Difference* measures how much each variable boosts the protected-class score (higher scores are worse for the applicant). It is calcu-

Figure 4: Univariate Test

		— Scorecard 1 —		— Scorecard 2 —	
Variable	Class Means, Difference	Point Weight	Point Difference	Point Weight	Point Difference
Fico	-45.00	-0.058	2.61	-0.045	2.03
Debt Ratio	2.00	0.635	1.27	0.899	1.80
Income (00,000)	-0.34	-1.212	0.41	-0.052	0.02
Mean Score			51.05		51.17

lated as the difference in means across the protected and non-protected classes multiplied by the point weight. In Scorecard 1, the point differences are fairly large for FICO and debt ratio. Comparing them to the mean score of 51.05 for the non-protected class, we see that FICO is responsible for a 5.11% increase in protected-class score and debt ratio is responsible for a 2.49% increase. Given the magnitude of these effects, both variables would generally be considered as candidates for disparate impact and corrective action would have to be investigated. (Income is responsible for only a 0.81% increase in score and would generally not be a candidate for disparate impact.) Because the potential for disparate impact was found to exist, the conclusion is that $\Psi > 0$ for Scorecard 1.

In Scorecard 2, the point differences for FICO and debt ratio are also large, with FICO being responsible for a 3.96% increase in score (based on a mean score of 51.17 for the non-protected class) and debt ratio being responsible for a 3.51%

increase. Again, both variables would generally be considered as candidates for disparate impact, indicating a conclusion of $\Psi > 0$ for Scorecard 2. This is a false positive finding of disparate impact when none exists.

The fact that the univariate test points to $\Psi > 0$ for Scorecard 2 illustrates a broader property of the univariate test — it tends to see *every* scorecard as having disparate impact potential. This is because the univariate test is largely driven by cross-class differences in scorecard-variable means, which is almost always sizable but which is unfortunately not related to the value of Ψ.

The multivariate disparate impact test consists of re-estimating the model with a protected-class indicator included as an explanatory variable. If the coefficient on the protected-class indicator is significant and positive, then the potential for disparate impact exists ($\Psi > 0$). If the coefficient is negative or not significant, then $\Psi = 0$.

Reestimating the models with a protected-class indicator included we got the results presented in Figure 5. Based on the significance of the protected class indicator, we conclude $\Psi > 0$ for Scorecard 1 and $\Psi = 0$ for Scorecard 2.

By looking at the change in the estimated coefficients relative to the original equations, the multivariate test lets us identify which variables are driving the disparate impact . The change in the coefficient on FICO is -56%, on debt ratio it is 190%, and on income it is 100%, indicating that the variables are all potential drivers of disparate impact. Note that the ordering of the variables in terms of importance is different than in the univariate case. From highest potential for disparate impact to lowest potential, the current ordering is 1) debt ratio,

Figure 5: Multivariate Test - Estimated Equation with Protected Class Indicator

	— Scorecard 1 —		— Scorecard 2 —	
Variable	Estimated Coefficient	Significance	Estimated Coefficient	Significance
Intercept	-10.33	<.0001	-10.52	<.0001
Fico	-0.025	<.0001	-0.026	<.0001
Debt Ratio	0.510	<.0001	0.519	<.0001
Income (00,000)	-0.003	0.9116	-0.028	.3708
Protected	10.27	<.0001	0.04	.6436

2) income, 3) FICO, while the ordering in the univariate approach is 1) FICO, 2) debt ratio, 3) income. In general, such re-orderings can be expected because the multivariate approach takes account of co-variation among variables, which may amplify or attenuate the impact of any individual variable on the scores, and the univariate approach does not.

The usual solution for correcting disparate impact is to drop variables from the scorecard. (The decision to drop a variable is made by subjectively weighing the variable's business relevance against its contribution to the average score difference across classes.) We investigated this corrective action by dropping each variable in turn from the scorecard, and then recalculating the new value of Ψ. Figure 6 presents the results. The first column gives the value of Ψ that we get when we rescore the applications using the original scorecard coefficients but omitting the indicated variable. The second column gives the value of Ψ that we get when we re-estimate the model after dropping a variable and then use the new coefficients to score the applications. In every case, Ψ is substantially greater than zero. We conclude that dropping variables is not an effective corrective action.

Figure 6: Effect of Dropping Variables, Scorecard 1

Variable	Without Re-estimation	With Re-estimated
Fico	$\Psi = 12.1$	$\Psi = 12.4$
Debt Ratio	$\Psi = 83.1$	$\Psi = 83.3$
Income	$\Psi = 22.9$	$\Psi = 22.9$

6 A New Corrective Procedure

We have seen that the corrective action of dropping variables does not achieve the desired objective of $\Psi = 0$. The reason it fails is that disparate impact is due not to the mere presence of particular variables in the estimated model, but to the overall pattern of correlation among all the variables, including the error term and the protected class indicator. Correcting disparate impact requires a more drastic solution than dropping variables. We propose the solution suggested by equation (5). That is, we propose that modelers adopt the practice of including minority status as a control variable during model development.

Figure 7: Corrected Scorecard for Process 1

Variable	Point Weight
Base Points	70
Fico	-0.090
Debt Ratio	1.839
Income (00,000)	-
	Ψ=0.0

Effectively, this would amount to replacing the Scorecard 1 point weights with those from (5), appropriately transformed. The new scorecard would look like Figure 7. This scorecard achieves $\Psi = 0$.

7 Summary

We defined a measure of disparate impact, Ψ, as the percentage of rejected protected-class members that would not have been rejected using a class-neutral score. We then defined a 'good' disparate impact test as one that differentiates between $\Psi > 0$ and $\Psi = 0$, and a 'good' corrective action plan as one that achieves $\Psi = 0$. We set up two test scorecards, one in which $\Psi > 0$ (Scorecard 1) and one in which $\Psi = 0$ (Scorecard 2). We then evaluated two disparate impact testing procedures, univariate and multivariate. Both testing procedures identified Scorecard 1 as a case where $\Psi > 0$, however, only the multivariate test correctly identified Scorecard 2 as a case where $\Psi = 0$. (The fact that the univariate test falsely concluded $\Psi > 0$ for Scorecard 2 illustrates the 'false positive' problem that characterizes typical univariate tests for disparate impact.) Next we evaluated the standard corrective action of eliminating variables and showed that it does not actually correct disparate impact — the closest we got to $\Psi = 0$ was $\Psi = 12.1$

In the final section we presented a new corrective action of retaining the original set of variables but adjusting their scorecard weights. While this approach produced the desired result, it did require use of protected class status in model development, which may be illegal under ECOA. The approach we offer does achieve $\Psi = 0$

REFERENCES

Arrow, Kenneth "The Theory of Discrimination". 1973. In O. Aschenfelter and A. Reese (editors) Discrimination in Labor Markets, Princeton University Press.

Avery, Robert B., Raphael W. Bostic, Paul S. Calem, and Glenn B. Canner. 1996. "Credit Risk, Credit Scoring, and the Performance of Home Mortgages". Federal Reserve Bulletin 82: 621-648.

Avery, Robert B., Raphael W. Bostic, Paul S. Calem, and Glenn B. Canner. 2000. "Credit Scoring: Statistical Issues and Evidence from Credit Bureau Files" Real Estate Economics 28(3): 523-547.

Becker, Gary S. The Economics of Discrimination. University of Chicago Press, 1959.

Board of Governors of the Federal Reserve. 1999. Consumer and Community Affairs CA99-1. Washington, DC.

Collins, M. Cary, Keith Harvey, and Peter Nigro. 2001. "The Influence of Bureau Scores, Customized Scores and Judgmental Review on the Bank Underwriting Decision Making Process". Paper presented at the Federal Reserve System Changing Financial Markets & Community Development Conference, Washington, DC, April 5-6, 2001.

Degroot, Morris H., Stephen E. Feinberg, and Joseph B. Kadane. 1994. Statistics and the Law, John Wiley and Sons, New York.

Federal Financial Institutions Examination Council. January 1999. "Interagency Fair Lending Examination Procedures". Washington, DC.

Federal Financial Institutions Examination Council "Interagency Task Force on Fair Lending Policy Statement, March 8, 1994. Washington, DC.

Fix, Michael and Raymond Struyk. 1993. Clear and Convincing Evidence. The Urban Institute, Washington, D.C.

LaCour-Little, Michael. 1999. "Discrimination in Mortgage Lending: A Critical Review of the Literature". Journal of Real Estate Literature 7(1): 15-52.

Ladd, Helen F. 1998. "Evidence on Discrimination in Mortgage Lending". Journal of Economic Perspectives 12(2): 41-62.

Martell, Javier, Paul Panichelli, Rich Strauch, and Sally Taylor-Shoff. 1999. "The Effectiveness of Scoring on Low-to-Moderate-Income and High-Minority Area Populations". Fair, Isaac, and Company, San Rafael, California.

Mays, Elizabeth, editor. 2001. Handbook of Credit Scoring. Glenlake Publishing Company, Chicago.

Munnell, Alicia H., Lynn Browne, James McEneaney, and Geoffrey Tootell. 1996. "Mortgage Lending in Boston: Interpreting HMDA Data". 1996. American Economic Review 86 (1): 25-54.

Kaye, David "Statistical Evidence of Discrimination". 1982. Journal of the American Statistical Association 77: 773-783.

Kaye, David H. and Mikel Aickin (editors). 1986. Statistical Methods in Discrimination Litigation, Marcel Dekker, Inc., New York.

Phelps, Edmund S. "The Statistical Theory of Racism and Sexism". 1972. American Economic Review 62(4): 659-661.

Sandler, Andrew L. And Jonathan Biran: 1995. "The Improper Use of Statistics in Mortgage Lending Discrimination Actions". In Yezer, Anthony (editor) Fair Lending Analysis, American Bankers Association, Washington, DC.

Siskin, Bernard R. "Comparing the Role of Statistics in Lending and Employment Cases". 1995. In Yezer, Anthony (editor): Fair Lending Analysis, American Bankers Association, Washington, DC.

U.S. Department of Justice. 2000. Amicus Curiae Brief filed in Support of Plaintiffs' Opposition to Defendant's Motion for Summary Judgment, United States District Court, No. 3-98-0223, Nissan Motor Acceptance Corporation, Defendant.